

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein nucleic search, using frame_plus_p2n model

Run on: January 16, 2003, 16:55:57 ; Search time 968.471 Seconds
(without alignments)
330.553 Million cell updates/sec

Title: US-09-856-070-23
Perfect score: 55
Sequence: 1 ELMLRLQDYEE 11

Scoring table:
BL0SOM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
delop 6.0 , delext 7.0

Searched: 2054640 seqs, 14561402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+_p2n.m+del -PEV-zip
-Q-/cdn2_1/USPto_spool/US09856070/rnast_14012003_155844_1621/app_query_tas_a_1_1592
-DB-GenEmbl -QMT-fasap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=3
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human240.cdi -LISP=45
-DOCALLIGN=200 -THR_SCORE=100 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEP=LOCAL
-OUTMT-plc -NPM-ext -HCAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER-insors6070.arin_1_155844_1621/rnast_14012003_155844_1621 -NPM=6 -Tpm=3
-NO_XLPXY -NO_WMAP -LAPQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAUD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELTOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_bri.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_phi.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: cm_ba.*
- 16: cm_fun.*
- 17: cm_hum.*
- 18: cm_to.*
- 19: cm_mu.*
- 20: cm_m.*
- 21: cm_ov.*
- 22: cm_ov.*
- 23: cm_par.*
- 24: cm_ph.*
- 25: cm_pl.*
- 26: cm_ro.*
- 27: cm_sts.*
- 28: cm_un.*

- 29: cm_vl.*
- 30: cm_htg_hum.*
- 31: cm_htg_inv.*
- 32: cm_htg_other.*
- 33: cm_htg_mus.*
- 34: cm_htg_pln.*
- 35: cm_htg_ror.*
- 36: cm_htg_mam.*
- 37: cm_htg_vrt.*
- 38: cm_sy.*
- 39: cm_htgo_hum.*
- 40: cm_htgo_mus.*
- 41: cm_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	100.0	2514	4	NC012814 Bos taurus
2	55	100.0	2930	9	AF011074 Human cytov
3	55	100.0	3044	6	AX411074 Sequence
4	55	100.0	3044	6	AX411074 Sequence
5	55	100.0	3047	6	AX440476 Sequence
6	55	100.0	3072	9	U5802625 Homo sapi
7	55	100.0	3173	9	U5802625 Homo sapi
8	55	100.0	215705	9	AF189213 Homo sapi
9	55	94.5	478	9	AF189213 Homo sapi
10	52	94.5	2701	10	MMEZR
11	52	94.5	183362	2	AF087635 Mus muscu
12	52	94.5	207782	2	AF087635 Mus muscu
13	52	94.5	217917	2	AF087635 Mus muscu
14	52	94.5	286508	2	AF087635 Mus muscu
15	50	90.9	475	9	AF188897 Homo sapi
16	48	87.3	1445	10	AF450294
17	48	87.3	182233	2	AF113744
18	48	87.3	189873	2	AF113744
19	45	81.8	426	9	AF199015 Homo sapi
20	42	76.4	182756	2	AF004891 Gyrza sat
21	42	76.4	187419	2	AF005053 Gyrza sat
22	42	76.4	233738	9	AC007956 Homo sapi
23	41	74.5	136822	5	AC091091 Gallus ga
24	40	72.7	243	8	YSALK2B
25	40	72.7	2545	5	AF019790 Gallus ga
26	40	72.7	105596	2	AC099130 Rattus no
27	40	72.7	150928	2	AC099130 Rattus no
28	40	72.7	158179	2	U5802625 Gyrza sat
29	39	70.9	592	5	AX229998
30	39	70.9	1678	9	U5805994 Human clone
31	39	70.9	1696	9	U5805995 Human clone
32	39	70.9	1922	9	U5807408 Human clone
33	39	70.9	2075	9	AF096451 Homo sapi
34	39	70.9	2452	9	AF091361 Homo sapi
35	39	70.9	2784	9	AF095771 Homo sapi
36	39	70.9	2790	1	SACNTPASE
37	39	70.9	4875	1	SAC437617 Sulfolobu
38	39	70.9	6583	9	AF082523 Homo sapi
39	39	70.9	10079	7	AF155037 Ateromon
40	39	70.9	34831	3	CEC4762
41	39	70.9	50459	9	AF002451 Homo sapi
42	39	70.9	95047	2	AC097135 Rattus no
43	39	70.9	95560	2	AC127620 Rattus no
44	39	70.9	96879	9	HS384121 Human DNA
45	39	70.9	93489	2	AF084843 Homo sapi

ALIGNMENTS

RESULT 1

HOVEZRINA
 DEFINITION Bos taurus ezrin mRNA, complete cds.
 ACCESSION M98498
 VERSION M98498.1 GI:289407
 KEYWORDS ezrin.
 SOURCE Bos taurus brain cDNA to mRNA
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 2514)
 Bergson,C.M., Zhao,H., Saijoh,K., Human,R.S. and Nestler,E.J.
 Ezrin and osteonectin, two proteins associated with cell shape and growth, are enriched in the locus coeruleus
 Mol. Cell. Neurosci. 4, 64-73 (1993)
 JOURNAL
 FEATURES
 source Location/Qualifiers
 1..2514
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /issue_type="brain"
 152..1897
 /codon_start=1
 /product="ezrin"
 /protein_id="AAA30510.1"
 /db_xref="GI:289408"
 /translation="MPKPIINRVITMDAEFAIQNTTGTGQEDQVVKTIIGLREVWY
 FGLQVYDNNKGFPTWIKLQKVSQAEVRKESPIQKFKPAKFPEDVAEELIQDTQKLF
 ELQVKEGILSDIEIGYPPETAVLLGSYVAQAKFGDYNKELIKAGYLGSERLIPQVMDQ
 HKITRQWDRLOVHABHRGMKDSAMLYLKIAQDLEMYGYNFYFKNNKKGIDML
 GVALGINLYKADKLTPIKIPGFWSEIRNISFNQKKFVKIPDKKAPDFVFPAPRIKI
 NKRILQICMNHPIYMPKPPPTIPVQOMKAPPKHQKLEPQQLTEKKPPPTVFE
 RPEQMREKFEIMIPLADYEETKPAEKELSDIQLALKLEEFKKAQEEAGLEAD
 RLAAARAKFEIERQADQIKSQQLATLAETAKIALLEAPRPKNVEVWOLPAK
 EAQDDLKTRFELHVMTPAPPFPVFPVNVVHIEGPGUEGTELSAEISSEGLIDRN
 EFKRIIEAKNFRVQRQIMLISLSELSQARDENKKTNDIHNENMROGRUKYKTIROI
 KQNKQKQRIDEEAL"
 644 a 658 c 738 g 474 t
 BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0154 Length: 2514
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0
 US-09-856-070-23 (1-11) x HOVEZRINA (1-2514)
 QY 1 GluteuMetLeuArqLeuGlnAspTyrGluGlu 11
 Db 1187 GAGTGTGATGCTCGAGTTTCAGSACTAGSAGAA 1219
 |||||||||||||||||||||||||||||||||||
 RESULT 2
 HUMVIL2
 LOCUS Human cytovillin 2 (VIL2) mRNA, complete cds.
 DEFINITION J05021
 ACCESSION J05021.1 GI:340216
 VERSION cytovillin; cytovillin 2; microvillar protein.
 KEYWORDS Human placenta, cDNA to mRNA.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2430)
 Turunen,O., Winqvist,R., Pakkanen,R., Grzeschik,K.H., Wahlstrom,T.
 and Vaheri,A.
 Cytovillin, a microvillar Mr 75,000 protein, cDNA sequence,
 prokaryotic expression, and chromosomal localization
 J. Biol. Chem. 264 (28), 16727-16732 (1989)
 JOURNAL
 MEDLINE
 PUBMED 2674140

COMMENT Draft entry and computer-readable sequence for 11 kindly submitted
 by O.Turunen, 31-AUG-1989.
 FEATURES
 source Location/Qualifiers
 1..2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="6q22 q27"
 1..2930
 /gene="VIL2"
 110..1837
 /gene="VIL2"
 /note="cytovillin 2"
 /codon_start=1
 /protein_id="AAA612/8.1"
 /db_xref="GI:340217"
 /db_xref="GDB:600-120-489"
 /translation="MDAEFAIQNTTGTGQEDQVVKTIIGLREVWYFGLIYVNDKGF
 PIWKIDKKVSQAEVRKESPIQKFKPAKFPEDVAEELIQDTQKLF
 FLYCPETAVLLGSYVAQAKFGDYNKVKSYLSSEKLIQVMDQHKITRQWDR
 LOVHABHRGMKDSAMLYLKIAQDLEMYGYNFYFKNNKKTIDILQALGINLYE
 KDKLTPIKIPGFWSEIRNISFNQKKFVKIPDKKAPDFVFPAPRIKIRILQICMNH
 IELYMRRKPPDTIEVQOMKAPPEEKHQKLEPQQLTEKKARRETVEREQMREKE
 ELMRLQDTEETKKAARELSQIQRAIQLEFFERPAQEEAEARLEADPMALRAKEEL
 ERLAVTQIKSQQLAAELARYTAKIALLEAPRPKNVEVWOLPAK
 ERLVMTAPPPPPVFPVFPVNVVHIEGPGUEGTELSAEISSEGLIDRN
 EFKRIIEAKNFRVQRQIMLISLSELSQARDENKKTNDIHNENMROGRUKYKTIROI
 FAKNFRVQRQIMLISLSELSQARDENKKTNDIHNENMROGRUKYKTIROI
 QRIDEEAL"
 793 a 658 c 821 g 658 t
 BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0183 Length: 2930
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-856-070-23 (1-11) x HUMVIL2 (1-2930)
 QY 1 GluteuMetLeuArqLeuGlnAspTyrGluGlu 11
 Db 1112 GAGTGTGATGCTCGGCTGCAGGACTATGAGGAG 1144
 |||||||||||||||||||||||||||||||||||
 RESULT 3
 AX411074
 LOCUS AX411074 3044 bp DNA
 DEFINITION Sequence 3721 from Patent WO229103.
 ACCESSION AX411074
 VERSION AX411074.1 GI:21443779
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Alvaes,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
 Gene expression profiles in liver cancer
 Patent: WO 0229103-A 3721 11-APR 2002;
 GENE LOGIC INC (US)
 FEATURES
 source Location/Qualifiers
 1..3044
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="EMBL/GenBank Accession No. X51521"
 826 a 687 c 855 g 675 t
 BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0191 Length: 3044
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 6 Caps: 0

US-09-856-070-23 (1-11) x AX411074 (1-3044)

QY 1 GlutMetGluArgGluGlnAspTyrGluGlu 11
|||||
DB 1153 GAGTGGATGCTGGGCTGGAGGACTATGAGGAG 1185

RESULT 4

HSEZRIN
LOCUS Human mRNA for ezrin.
ACCESSION X51521
VERSION 1 GI:31282
KEYWORDS ezrin; kinase substrate; microvilli protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 3044)
AUTHORS Hunter, T.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1990) Hunter, T., The Salk Institute, Molecular
Biology and Virology Laboratory, 10610 North Torrey Pines Road, San
Diego, CA 92138, USA
REFERENCE 2 (Bases 1 to 3044)
AUTHORS Gould, K.L., Breitscher, A., Esch, F.S., and Hunter, T.
TITLE cDNA cloning and sequencing of the protein-tyrosine kinase
substrate, ezrin, reveals homology to band 4.1
JOURNAL EMBO J. 8 (13), 4133-4142 (1989)
MEDLINE 90076135
PUBMED 2591371
COMMENT See also <J05021>.

FEATURES

source
Location/Qualifiers
1..3044
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="F6"
/cell_line="HeLa"
/clone_lib="Okayama-Berq"
118..1878
/note="ezrin (AA 1-586)"
/codon_start=1
/protein_id="CAA35893.1"
/db_xref="GI:31283"
/db_xref="SWISS-PROT:P15311"
/translation="MKP INVRVIMIAELERPAIQNIFKGLFVGVKIEGLPEVWY
PELVYVINKPEFWKIKAFVSAGVFKNDIAQKFEAKYIEVAELIICLLQKKE
FLOVKKQLSLELYCPDEAVILGSAVQAKCDYNNKPKSKYLSSERLIDQRYMDQ
IKLTQWEDRIQVWIAEIRGMLKDNAMLEYLKIADQLEMYGINTFEIKNNKGLDML
NKRLQCMNHELYMPKPKPTIEVQMKAAQAFKIQKLEPQKLETERKPRETVE
REKQMKREKLEMLRIQYIEKIKAAKRELSPQIRALQLEKRAKRAEAKLEAD
KMAALRAKLELRQAVDQIKSQDQIAELAEYIAKLAILEARRKKEDVPEWQHRK
EAODLVKPELRIHVMIAAPPPIPPVYRPSVHYVQESLQDKCAEPTQSYAFSLSEGI
RDGNEKFKITEAKNEKVRQLVTLTSLSELNAKENKFIINDIHNENMKVQREKYK
TLQIRGNTKQRIDEEAL"
1709..1711
/note="ezrin was tgc in [1]"
/citation=[1]
old_sequence

BASE COUNT 826 a 687 c 855 g 675 t 1 others
ORIGIN

Alignment Scores:
Pred. No. 0.0191 Length: 3044
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-856-070-23 (1-11) x HSEZRIN (1-3044)

QY 1 GlutMetGluArgGluGlnAspTyrGluGlu 11
|||||
DB 1153 GAGTGGATGCTGGGCTGGAGGACTATGAGGAG 1185

RESULT 5

AX440476
LOCUS Sequence 329 from Patent WO0190154.
ACCESSION AX440476
VERSION 1 GI:21665286
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Xu, J., Mitcham, J.B., Harlocker, S.L., Dillon, D.C., Secriest, H.,
Ledes, M.J., Aljate, P.A., Fung, S.P., Mannion, J., Benson, D.P., and
Carter, D.
TITLE Compositions and methods for the therapy and diagnosis of ovarian
cancer
JOURNAL Patent: WO 0190154-A 329 29-NOV-2001;
CORIXA CORPORATION (US)
FEATURES location/Qualifiers
1..3047
/organism="Homo sapiens"
/db_xref="Laxon:9606"

BASE COUNT 828 a 687 c 856 g 675 t 1 others
ORIGIN

Alignment Scores:
Pred. No. 0.0191 Length: 3047
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-856-070-23 (1-11) x AX440476 (1-3047)

QY 1 GlutMetGluArgGluGlnAspTyrGluGlu 11
|||||
DB 1153 GAGTGGATGCTGGGCTGGAGGACTATGAGGAG 1185

RESULT 6

HSM802625
LOCUS Homo sapiens mRNA, cDNA (AF4762157) (from clone AF4762157);
DEFINITION complete cds.
ACCESSION AF152086
VERSION 1 GI:7328174
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 3072)
AUTHORS Ottenwaeider, B., Obermaier, B., Mewes, H.W., Weil, B., and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the cDNA
sequencing Consortium of the German Genome Project. This clone
(AF4762157) is available at the RZPD in Berlin. Please contact
the RZPD, Rossourcezentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
source
1..3072

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6q22-q27"
/clone="DKFZp762H157"
/tissue_type="melanoma (McMv cell line)"
/clone_lib="762 (SYMsym; hmc12). Vector pSpcrt1, host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..1894
/clone="DKFZp762H157"
1..1894
/clone="DKFZp762H157"
/notes="ezrin"
/product="hypothetical protein"
/protein_id="CAH82418.1"
/db_xref="GI:7428175"
/translacion="GPGVGRVVAAGSTSTAGAGAGQPIPCWIVVPSGGIEDTIQKPK
MPKPINRVVTMTAELEFAIQNTTIGKQVVKTIIGLRVWYFGLHYVDNKGFTW
LKLDKRVSAQVFFENPLQAFKFAKFFVFAEELIADITKLIFFLAVKEGILSDEIV
CPDEFIALLGSAVQAKGIVANKFVKKSYI SSPII PGPVMIQHKIIPDGMFQIOW
WHAHFGMIKDNAMILEYLIKADLEMYGIVNVEIKNKKGTIDIWGVNAGINIVKTD
KLPKIGIPWSELRNISENKKFVILKPIUKAAPHVYAPRLRNKRILOLCMNHEL
YMRKRPDIIVQVQKKAQAKBEKHQKLERQLETEKKRRETVEREKOMRREELM
LRQDTEKKKAERELSPQIOPALQLEFEKRPAGQFAPFLEADPMALPAKRELEPQ
AVDQTSQGLAAFIATYAKTALILEFAFPKDEVEEWHEAKAODLVKKEELH
LVMTAPPPPPVYEFVSYHVQESIQDAGAPFIQYSALISSEGIKIDRNEKRIIFAE
KNKRVQRIILISSELISQARDENKRTNDIHNENMRQCRDKYKTLIQIKQCNIKQRI
DEFEAL"
3032..3037
polyA_signal
polyA_site
3051..3051
BASE COUNT 846 a 688 c 868 g 670 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0193 Length: 3072
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-856-070-23 (1-11) x USM802625 (1-3072)
QY 1 GlutLeuMetLeuArgLeuGlnAspTyrGluGlu 11
|||||
Db 1169 GAGTTGATGCTGGCGGCGGAGGACTATGAGGAG 1201
RESULT 7
LOCUS BC013903 3173 bp mRNA linear PRI 10-SEP-2001
DEFINITION Homo sapiens, Similar to villin 2 (ezrin), clone MCC:1584
IMAGE:2959499, mRNA, complete cds.
ACCESSION BC013903.1 GI:15530242
VERSION M87.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3173)
AUTHORS Strausberg,K.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A04, Bethesda, MD 20892-2540,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
CONTACT: MCC help desk
EMAIL: qcaphs@mail.nih.gov
Tissue Procurement: AICC
cDNA Library Preparation: Pabin laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahcy, Stephanie Ford, Julia
Greene, Mark Kettman and Anuradha Madan
Clone distribution: Mcv clone distribution information can be found
through the I.M.A.G.E. Consortium/Link site: http://image.llnl.gov
Series: IRAL Plate: 2 Row: a Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9257254.
FEATURES
source
1..3173
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MCC:1584 IMAGE:2959499"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B-R"
/vector="Vector, pOTB7"
142..1902
/codon_start=1
/product="Similar to villin 2 (ezrin)"
/protein_id="AAH13903.1"
/db_xref="GI:15530243"
/translacion="MPKPINRVVTMTAELEFAIQNTTIGKQVVKTIIGLRVWY
FGLHYVDNKGFTWIKLQKRVSAQVFFENPLQAFKFAKFFVFAEELIADITKLIFF
FLOVKGILSDRIYCPDEFIALLGSAVQAKGIVANKFVKKSYI SSPII PGPVMIQHKIIPDGMFQIOW
WHAHFGMIKDNAMILEYLIKADLEMYGIVNVEIKNKKGTIDIWGVNAGINIVKTD
KLPKIGIPWSELRNISENKKFVILKPIUKAAPHVYAPRLRNKRILOLCMNHEL
YMRKRPDIIVQVQKKAQAKBEKHQKLERQLETEKKRRETVEREKOMRREELM
LRQDTEKKKAERELSPQIOPALQLEFEKRPAGQFAPFLEADPMALPAKRELEPQ
AVDQTSQGLAAFIATYAKTALILEFAFPKDEVEEWHEAKAODLVKKEELH
LVMTAPPPPPVYEFVSYHVQESIQDAGAPFIQYSALISSEGIKIDRNEKRIIFAE
KNKRVQRIILISSELISQARDENKRTNDIHNENMRQCRDKYKTLIQIKQCNIKQRI
DEFEAL"
930 a 691 c 868 g 684 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 0.02 Length: 3173
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-856-070-23 (1-11) x BC013903 (1-3173)
QY 1 GlutLeuMetLeuArgLeuGlnAspTyrGluGlu 11
|||||
Db 1177 GAGTTGATGCTGGCGGCGGAGGACTATGAGGAG 1209
RESULT 8
LOCUS AL589931/c 215705 bp DNA linear PRI 15-SEP-2001
DEFINITION Human DNA sequence from clone Pp11-507C10 on Chromosome 6q25.2-26,
complete sequence.
ACCESSION AL589931 AC024785
VERSION AL589931.14 GI:15591370
KEYWORDS HIG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 215705)
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonesrequest@sanger.ac.uk
on Sep 12, 2001 this sequence version replaced gi:14586248.
During sequence assembly data is compared from overlapping clones.

```



```

misc_feature 20273..21432
              /note="assembly_name:Contig205"
misc_feature 21433..21573
              /note="assembly_name:Contig204"
misc_feature 21574..21580
              /note="assembly_name:Contig203"
misc_feature 21581..21585
              /note="assembly_name:Contig202"
misc_feature 21586..21797
              /note="assembly_name:Contig201"
BASE COUNT 5958 a 52987 c 45774 g 55341 t 207 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 11.1 Length: 21797
Score: 52.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 94.55% Indels: 0
DB: 2 Gaps: 0

```

US-09-856-070-23 (1-11) x A0223294 (1-21797)

QY 1 GluLeuMetLeuArgLeuGlnAspTyrGluGlu 11

Db 10729 GAGTATATTTGGATGATAGGA-TAGAGATG 10697

```

RESULT 14
AC125143 286508 bp DNA Linear HTG: OR AUG 2002
LOCUS Mus musculus chromosome UNK clone PP24-342H13, WAPPING DRAFT
DEFINITION
SEQUENCE, 10 unordered pieces.
AC125143
VERSION AC125143.1 GI:22138589
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.

```

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 286508)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 286508)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

```

```

COMMENT
----- Genome Center
Center, Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information
Center project name: M.BB0342H13
----- Summary Statistics

```

```

Sequencing vector: M13; 0%
Sequencing strategy: plasmid; 100%
Chemistry: Dye-terminator ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: phrap, version 0.99014
Consensus quality: 283924 bases at least Q40
Consensus quality: 285778 bases at least Q30
Consensus quality: 286826 bases at least Q20
Insert size: 255000, average 15
Insert size: 304647, sum-of-contigs
Quality coverage: 14.29 in Q20 bases; average 15
Quality coverage: 8.54 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.
1 1503: contig of 1503 bp in length
* 1504: gap of unknown length
* 1604: contig of 2982 bp in length
* 4585: gap of unknown length
* 4586: contig of 18201 bp in length
* 22887: gap of unknown length
* 22887: contig of 21961 bp in length
* 44947: gap of unknown length
* 45048: contig of 31906 bp in length
* 76954: gap of unknown length
* 77054: contig of 26715 bp in length
* 103769: gap of unknown length
* 103769: contig of 56889 bp in length
* 160757: gap of unknown length
* 160858: contig of 123560 bp in length
* 24417: gap of unknown length
* 24417: contig of 851 bp in length
* 284518: gap of unknown length
* 285369: contig of 1040 bp in length.
* 285469: Location/qualifiers
1..286508
/organism="Mus musculus"
/db_xref="taxon:10090"
/chr_nm="chr.m.10090"
/chr_nm="PP24-342H13"
1..1503
/note="assembly_name:Contig99"
1604..4585
/note="assembly_name:Contig100"
4586..22886
/note="assembly_name:Contig101"
22887..44947
/note="assembly_name:Contig102"
45048..76953
/note="assembly_name:Contig103"
77054..103768
/note="assembly_name:Contig104"
103769..160757
/note="assembly_name:Contig105"
160858..284417
/note="assembly_name:Contig106"
284518..285368
/note="assembly_name:Contig107"
285369..286508
/note="assembly_name:Contig108"
286509..286508
/note="assembly_name:Contig109"
77032 t 907 others

```

FEATURES

SOURCE

```

misc_feature 1..1503
misc_feature 1604..4585
misc_feature 4586..22886
misc_feature 22887..44947
misc_feature 45048..76953
misc_feature 77054..103768
misc_feature 103769..160757
misc_feature 160858..284417
misc_feature 284518..285368
misc_feature 285369..286508
misc_feature 286509..286508

```

BASE COUNT 78351 a 64449 c 65769 g 77032 t 907 others

ORIGIN

Alignment Scores:

```

Pred. No.: 15.1 Length: 286508
Score: 52.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 94.55% Indels: 0
DB: 2 Gaps: 0

```

US-09-856-070-23 (1-11) x A0223294 (1-286508)

QY 1 GluLeuMetLeuArgLeuGlnAspTyrGluGlu 11

Db 10729 GAGTATATTTGGATGATAGGA-TAGAGATG 10697

RESULT 15

AF188897

LOCUS AF188897

DEFINITION Homo sapiens ovary early mRNA, partial cds.

ACCESSION AF188897

PR1 19-OCT-1999

```

VERSION      AF188897.1  GI:6063144
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
              Homo sapiens
REFERENCE
AUTHORS      Chen,Z.C., Fadiei,A. and Naftolin,F.
TITLE        Ezrin gene mutation in ovarian cancer
JOURNAL      Unpublished
              2. (bases 1 to 475)
REFERENCE
AUTHORS      Chen,Z.C., Fadiei,A. and Naftolin,F.
TITLE        Direct Submission
JOURNAL      Submitted (22 SEP 1999) DB/GYN, Yale University, 333 Cedar Street,
              New Haven, CT 06511, USA
FEATURES
  source
    1..475
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="6"
      /map="6q22-q27"
      /cell_line="Bix1"
      /cell_type="epithelial cancer cells"
      /tissue_type="ovary"
    <1..475
      /note="villin 2: Kinase substrate"
      /codon_start=2
      /product="ezrin"
      /protein_id="MAF03155.1"
      /db_xref="GI:6063145"
      /translation="TGMVGINVFYVKNKKGIDILGVIALGLNIYKDDKLXPKIGPP
      WSEIKNISFNKKFVINDIKAPDFVYAFGLKINKKIDQLCMGNHLYMKKKKPD
      IEVQMKAFAPEEKIQLEPQILETKKPEXVERPKFQMPREKEELMLPLQDYE"
BASE COUNT   154 a   90 c   136 q   93 t   2 others
ORIGIN

Alignment Scores:
  Prod. No.:      0.0341      Length:      475
  Score:          50.00      Matches:      10
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:    90.91%      Indels:      0
  DB:              9          Gaps:      0

US 09-856-070-23 (1 11) x AF188897 (1-475)
QY      1 GluLeuMetIeuArqLeuGlnAspTyrGlu 10
      |||
DB      446 GACTTCATCTTCGGCGTCAGCACATCAAA 475

```

Search completed: January 16, 2003, 19:05:54
 Job time : 10:4.47 secs